SIMULATION TOOLS FOR STATISTICAL MODEL COMPARISON: AN APPLICATION TO UNOBSERVED COMPONENT MODELS VERSUS DYNAMIC REGRESSION MODELS

Bahar Dadashova
Blanca Arenas Ramírez
José Manuel Mira McWilliams
Francisco Aparicio Izquierdo

Universidad Politécnica de Madrid

Instituto Universitario de Investigación del Automóvil (INSIA) – UPM

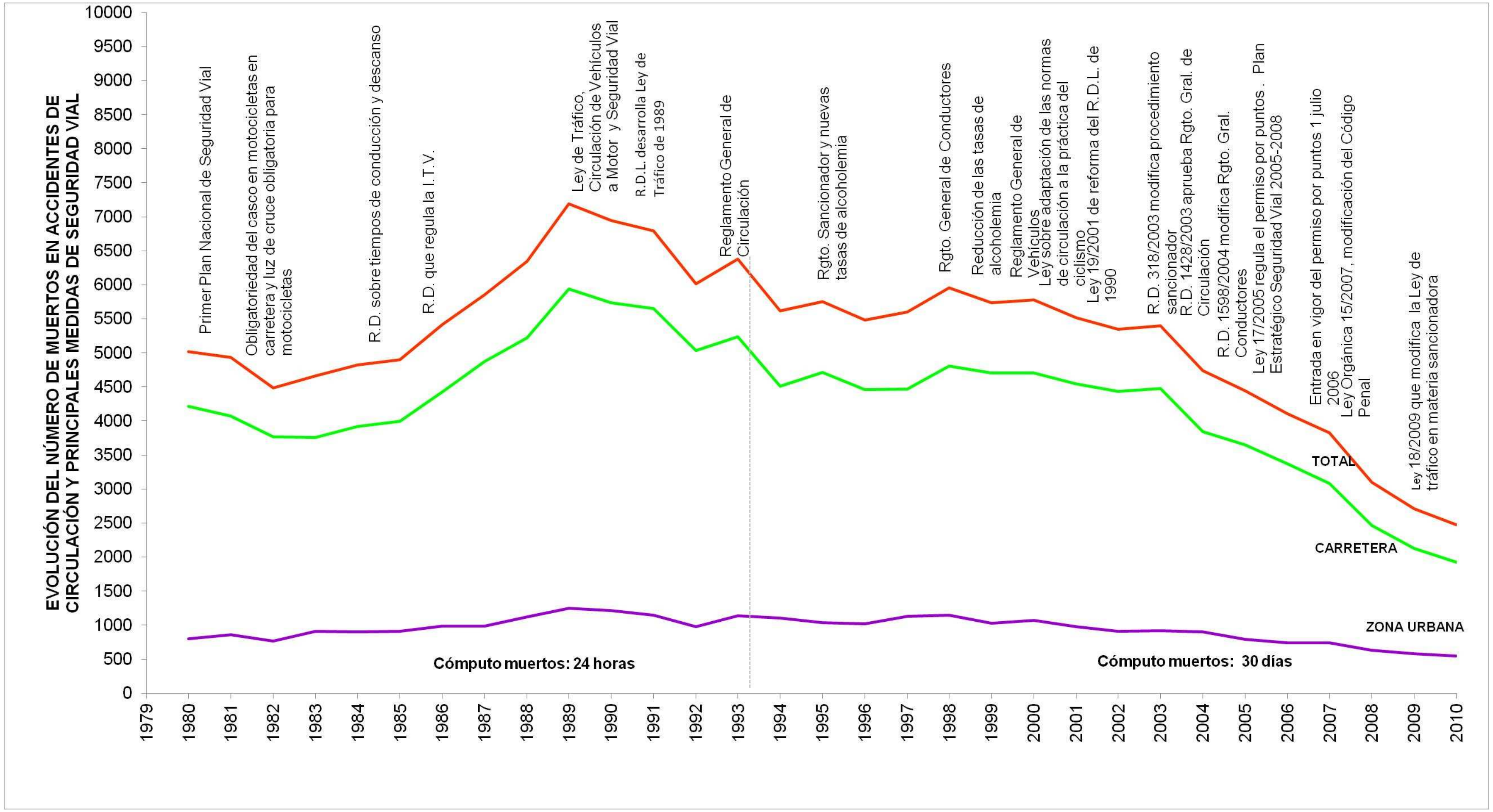
Centro de Investigación del Transporte (TRANSYT) – UPM

Instituto de Estudios de Automoción (IEA)

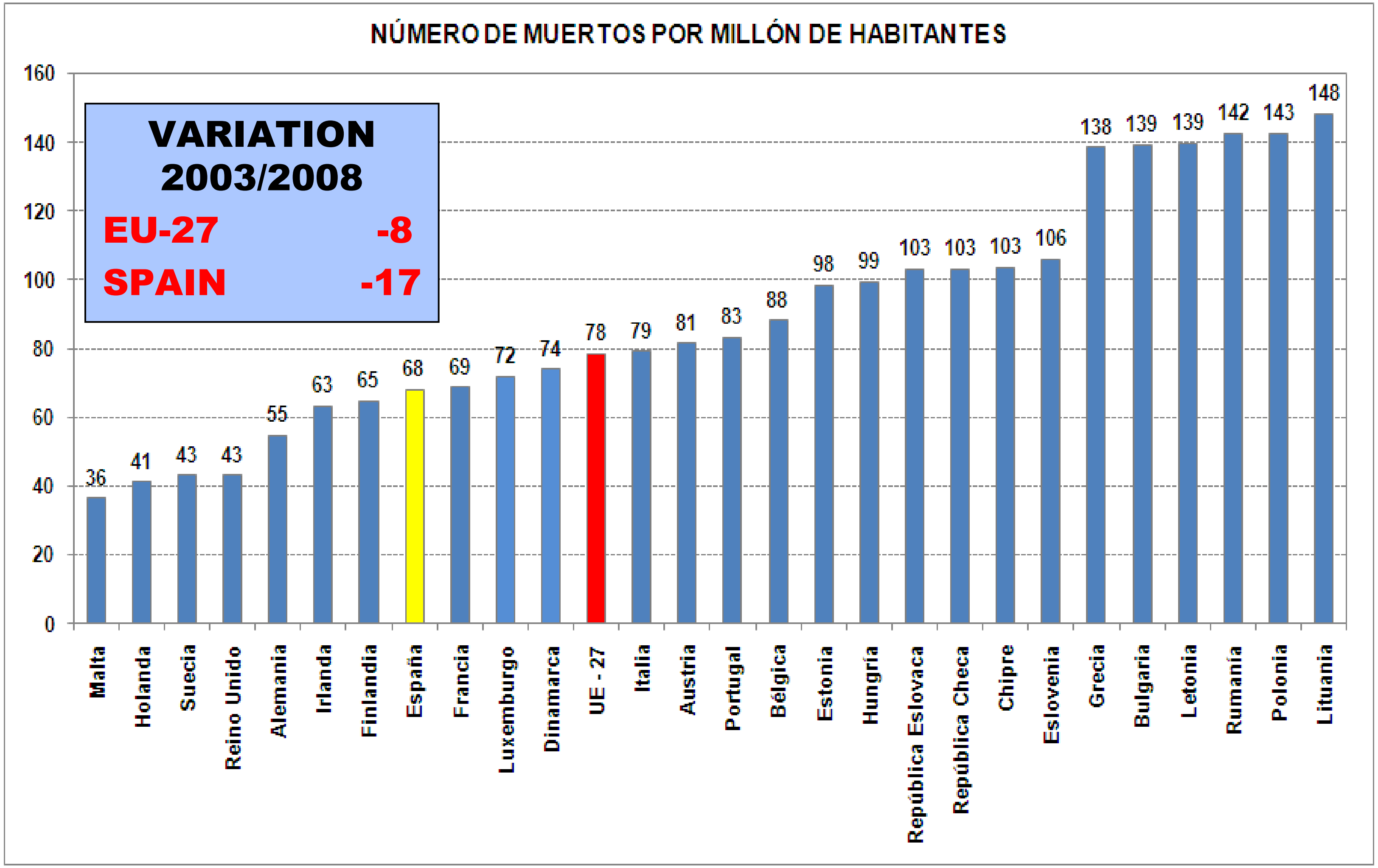
Universidad Carlos III de Madrid

ROAD ACCIDENTS IN SPAIN

FATAL ITY RATE IN SPANISH ROADS-LEGISLATIVE CHANGES



ROAD ACCIDENTS IN SPAIN



DYNAMIC MACRO MODELS FOR ROAD ACCIDENT ANALYSIS

- Tool to quantify the effects mentioned above mainly two time series model are used:
  - *Demand for road use, accidents and their gravity* (DRAG), developed by Gaudry (1984) and Gaudry and Lassarre (2000) ;
  - *Unobserved Components Models* (UCM) with intervention proposed by Harvey and Durbin (1986).

- *Main differences*
- UCM includes unobserved specific terms for *trend* and *seasonality*
- State equations for both terms
- UCM more complex and general
- DRAG has simpler interpretation

Question:

Relationship between two models, which DRAG terms capture UC trend and seasonality?

PURPOSE

- Better understanding of relationship between the two “competing” models
- We suppose UC is the true model and:

  - ✓ See how the DRAG parameter estimates capture the UC terms.
  - ✓ Relationship between parameters of both models.
  - ✓ Eventually effect on prediction as well.

STAGES OF COMPUTATIONAL EXPERIMENT

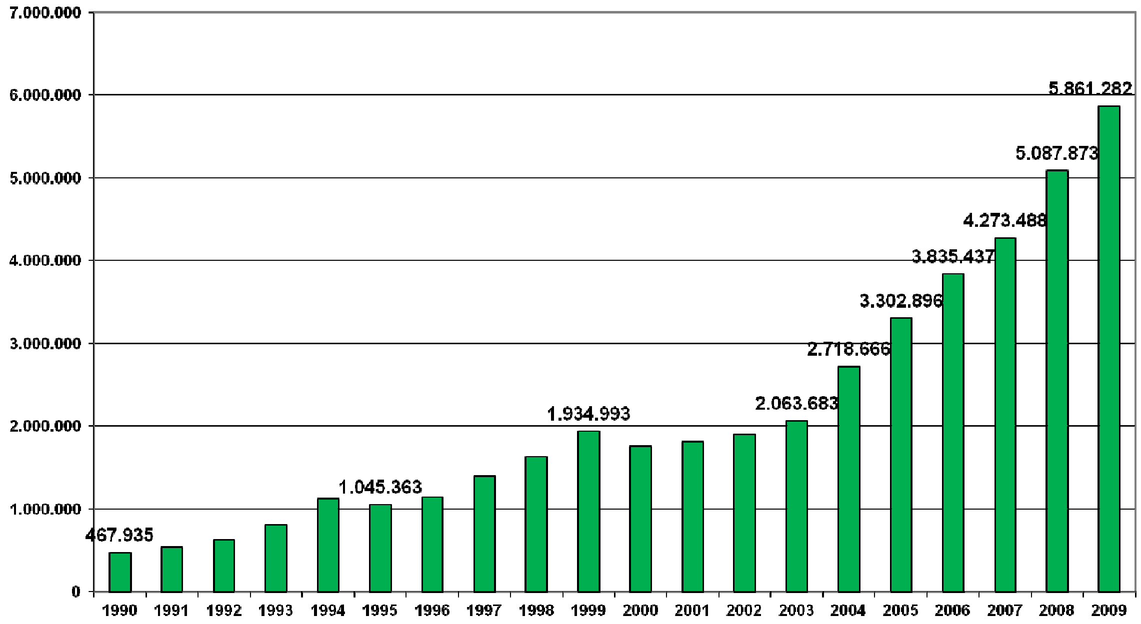
- The experiment is a **simulation** study with the following steps:
  - ✓ *Designing the simulations of the UC model using the results of the empirical study*
  - ✓ *Generation of UC samples (time series)*
  - ✓ *DRAG estimation:*
  - ✓ *ANOVA-type analysis of results*

INPUT VARIABLES

Table 1. Input variables

VARIABLES		
Exposure	Total fuel consumption	***COMTOT***
Economic factors	Unemployment rate in service sector	***PARSER***
Driver behavior	Number of alcohol control	***CONALC***
	Driver licensce suspencion	***SUSP***
Labor conditions	Number of labor days	***DLAB***

Number of alcohol controls (***CONALC***), **1990-2009**



STAGES OF COMPUTATIONAL EXPERIMENT: UC MODEL TO BE SIMULATED

The state space model consists of :

Mesurement equation :

$$y_t = \mu_t + \sum_{j=1}^{k} \delta_j x_{jt} + \omega_t + \varepsilon_t$$

State equation :

$$\mu_t = \mu_{t-1} + \eta_t$$

where ε_t and η_t are i.i.d $(0, \sigma_{\varepsilon_t})$ and $(0, \sigma_{\eta_t})$ respectively. And ω_t is an intervention variable

$$\omega_t = \begin{cases} 1 \text{ if } t < \tau \\ 0 \text{ if } t \geq \tau \end{cases}$$

Table 2. Input factors and the UCM estimators obtained from empiriacal work

UCM PARAMETERS –INPUT FACTORS		ESTIMATORS (*t*-value)
REGRESSORS	δ_{PARSER}	**0.97 (4.3)**
	δ_{CONALC}	**-0.24 (-2.43)**
	δ_{SUSP}	**-0.38 (-6.83)***
	δ_{DLAB}	**-0.31 (-2.63)**
	δ_{COMTOT}	**2.03 (4.3)**
TC VARIANCE	σ_{η_t}	**0.02**

Simulation model using the parameters in Table 2.:

$$y_t = \mu_t + \delta_{PARSER}\, x_{1t} + \delta_{CONALC}\, x_{2t} + \delta_{SUSP}\, x_{3t} + \delta_{DLAB}\, x_{4t} + \\ + \delta_{COMTOT}\, x_{5t} + \omega_{LCP} + \varepsilon_t$$

$$\omega_t = \begin{cases} 1 \text{ if } t < 96 \\ 0 \text{ if } t \geq 96 \end{cases}$$

[Go back](#)

STAGES OF COMPUTATIONAL EXPERIMENT: DRAG MODEL TO BE ESTIMATED

The general DRAG model is specified as follows :

$$Y_t^{(\lambda_y)} = \sum_{k=1}^{K} \beta_k X_t^{(\lambda_{xk})} + u_t$$

where u_t follows and $AR(l)$ process :

$$u_t = \sum_{l=1}^{r} \rho_l \, u_{t-l} + w_t$$

and the model variables are BCT :

$$Y_t^{(\lambda_y)} = \begin{cases} \dfrac{Y_t^{\lambda_y} - 1}{\lambda_y}, \text{if } \lambda_y \neq 0 \\ \ln(Y_t), \text{if } \lambda_y = 0 \end{cases}$$

Table 3. Response parameters

DRAG PARAMETERS	
REGRESSORS	β_{PARSER}
	β_{CONALC}
	β_{SUSP}
	β_{DLAB}
	β_{COMTOT}
AR PARAMETERS	ρ_1
	ρ_2
BCT COEFFICIENT	λ

- ❑ The DRAG model is estimated using TRIO which is devoloped by Lassarre and Gaudry ()
- ❑ DRAG parameters were estimated using :
  - ❑ Same independent variables as in UC;
  - ❑ Errors follow an autoregressive model of order 2.
- ❑ 8 response parameters in total

Go back

STAGES OF COMPUTATIONAL EXPERIMENT: EXPERIMENTAL DESIGN

Figure1: Experimental design

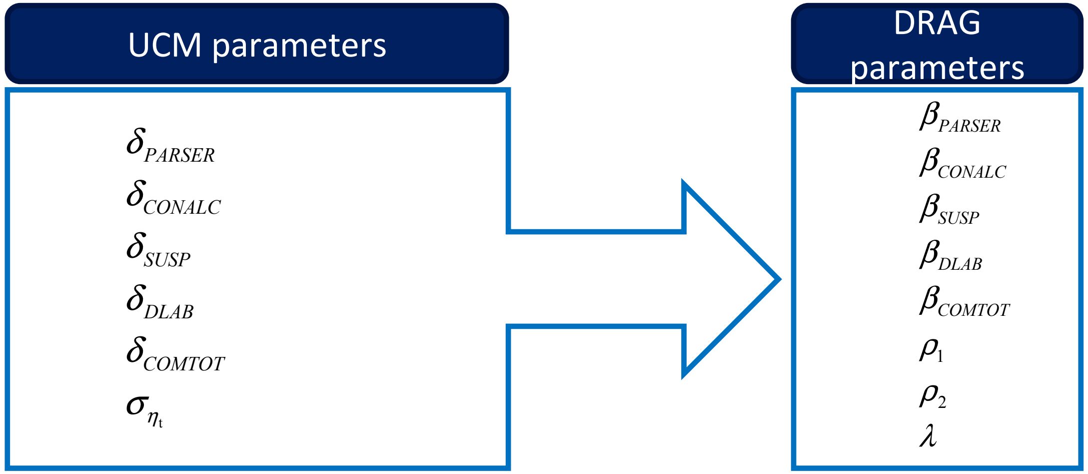


Table 4. Design matrix "+" and "–" for high and low levels of variables.

	δ_{PARSER}	δ_{CONALC}	δ_{SUSP}	δ_{DLAB}	…	
o	-	-	-	-	…	-0.341
δ_{PARSER}	+	-	-	-	…	-0.341
δ_{CONALC}	-	+	-	-	…	-0.306
δ_{SUSP}	-	-	+	-	…	-0.341
δ_{DLAB}	-	-	-	+	…	-0.341
…	…	…	…	…	…	…

STAGES OF COMPUTATIONAL EXPERIMENT: ANOVA ANALYSIS

- Estimation of main effects and interactions (2^{nd}, 3^{rd}, 4^{th}, 5^{th} and 6^{th} order) with the Yates algorithm:

$$\overline{\delta_1} = \frac{1}{4n}\left(-1 + \delta_1 - \delta_2 - \delta_3 - \delta_4 - \ldots\right)$$

$$\overline{\delta_1\delta_2} = \frac{1}{4n}\left(1 - \delta_1 - \delta_2 + \delta_3 + \ldots + \delta_1\delta_2 + \ldots\right)$$

- Computation of ANOVA - sums of squares

$$SS_{\delta_1} = \frac{1}{8n}\left(-1 + \delta_1 - \delta_2 - \delta_3 - \delta_4 + \ldots\right)^2$$

$$SS_{\delta_1\delta_2} = \frac{1}{8n}\left(1 - \delta_1 - \delta_2 + \delta_3 + \ldots + \delta_1\delta_2 + \ldots\right)^2$$

- ANOVA F-tests were applied to test their significance.
- We neglect interactions of 5^{th} and 6^{th} order, to estimate the error variance.

RESULTS AND INTERPRETATION

Table 5. Experimental design results for β_{CONALC} (considering up to 4th interaction effects only)

Variable	Effect estimate	Sum of squares	DF	Mean square	*F*-test	*P*-value
σ_{η_t}	0.1399	0.3133	1	0.3133	72.88889	<0.0001
$\delta_1\delta_{CONALC}\ \delta_3\delta_5$	-0.1216	0.2364	1	0.2364	55.00376	<0.0001
δ_{CONALC}	0.0919	0.1351	1	0.1351	31.41867	0.0001
…	…	…	…	…	…	…
Error	0.0301		7			
Total	1.4820		63			
R^2	0.9797					

- This model explains 97% of the variability in the total model, *F*-test is significan.
- For response β_{CONALC} ,the most significant effects and interactions:
  - ✓ Trend variance, (σ_{η_t})
  - ✓ Second UCM regression coefficient, (δ_{CONALC})

CONCLUSIONS

- Results are as expected intuitively, the trend effect has more complex interpretation. The DRAG regression coefficient is capturing :
  - ✓ the change in the corresponding UCM coefficient,
  - ✓ as well as the change in the variance of the trend component!

- Future research:
  - Effect on prediction errors
  - MANOVA instead of individual ANOVAs
  - Regression models including only significant effects and interactions.
  - In deterministic versus stochastic trend analysis

REFERENCES

- **Aparicio Izquierdo, F., Arenas Ramírez, B., Bernardos Rodríguez, E. (2011 a).** The interurban Drag-Spain model: The main influential factors of road accidents in Spain, *Research in Transportation Economics.* doi:10.1016/j.retrec.2011.08.011.

- **Bijleveld, F., Commandeur, Gould, Ph. and J., Koopman, S. J. (2008).** Model-based measurement of latent risk in time series with applications*. Journal of the Royal Statistical Society: Series A,* Vol. 171, Iss. 1. pp. 265–277.

- **Bijleveld, F., Commandeur, J., Koopman, S. J., and Van Montfort K. (2010).** Multivariate non-linear time series modeling of exposure and risk in road safety research. *Journal of the Royal Statistical Society,* Vol. 59, part 1, pp. 145-161.

- **Gaudry, M. and Lassarre, S. (2000).** Structural road accident models. The international DRAG family. *Elsevier Science.*

- **Harvey, A. C. and Durbin, J. (1986)**. The effects of seat belt legislation on British road causalities: a case study in structural time series modeling. *Journal of the Royal Statistical Society: Series A*, vol. 149, pp. 187-227.

- **Hermans, E., Wets, G., and Van den Bossche, F. (2005).** Describing the evolution in the number of highway deaths by decomposition in exposure, accident risk, and fatality risk. *Transportation Research Institute, Hasselt University.*

- **Hermans, E., Wets, G., and Van den Bossche, F. (2006).** Frequency and severity of Belgian road traffic accidents studied by state-space methods. *Journal of Transportation and Statistics*, 9(1), pp. 63-76.

- **Montgomery, D. C. (2000)**. Design and analysis of experiments. Fifth edition. *John Wiley & Sons, New York, NY.*

- **Peña, D. (1987).** Estadística. Modelos y métodos. 2. Modelos lineales y series temporales. *Alianza editorial.*

THANK YOU FOR YOUR ATTENTION!

seio 2012

XXXIII Congreso Nacional de Estadística e Investigación Operativa y de las VII Jornadas de Estadística Pública